The effect of task on visual attention in interactive virtual environments

JACOB HADNETT-HUNTER, University of Bath
GEORGE NICOLAOU, University of Bath
EAMONN O'NEILL, University of Bath
MICHAEL PROULX, University of Bath

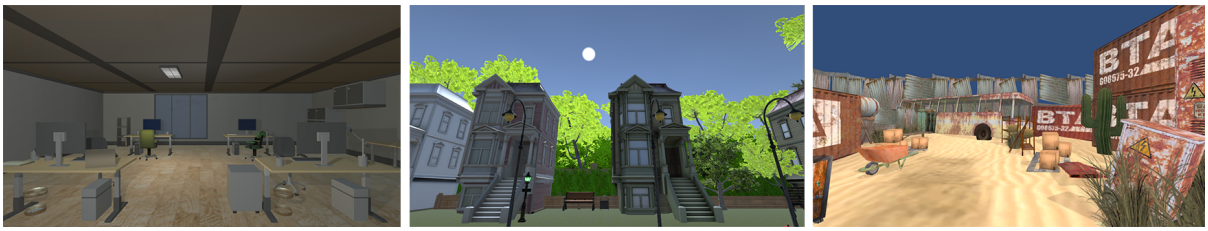

Fig. 1. The three virtual environments in which participants performed freeview, search and navigation tasks. On the left is an indoor office space. In the middle, a suburban street. On the right is a desert junkyard.

Virtual environments for gaming and simulation provide dynamic and adaptive experiences but, despite advances in multi-sensory interfaces, these are still primarily visual experiences. In order to support real time dynamic adaptation, interactive virtual environments could implement techniques to predict and manipulate human visual attention. One promising way of developing such techniques is to base them on psychophysical observations, an approach that requires a sound understanding of visual attention allocation. Understanding how this allocation of visual attention changes depending on a user's task offers clear benefits in developing these techniques and improving virtual environment design. With this aim, we investigated the effect of task on visual attention in interactive virtual environments. We recorded fixation data from participants completing freeview, search and navigation tasks in three different virtual environments. We quantified visual attention differences between conditions by identifying the predictiveness of a low-level saliency model, and its corresponding color, intensity and orientation feature conspicuity maps, as well as measuring fixation center bias, depth and duration as well as saccade amplitude. Our results show that task does affect visual attention in virtual environments. Navigation relies more than search or freeview on intensity conspicuity to allocate visual attention. Navigation also produces fixations that are more central, longer, and deeper into the scenes. Further, our results suggest that it is difficult to distinguish between freeview and search tasks. These results provide important guidance for designing virtual environments for human interaction, as well as identifying future avenues of research for developing 'attention-aware' virtual worlds.

Authors' addresses: Jacob Hadnett-Hunter, J.M.Elliott.Hadnett-Hunter@bath.ac.uk, University of Bath, Claverton Down, Bath, Somerset, BA2 7AY; George Nicolaou, University of Bath, Claverton Down, Bath, Somerset, BA2 7AY; Eamonn O'Neill, E.ONeill@bath.ac.uk, University of Bath, Claverton Down, Bath, Somerset, BA2 7AY; Michael Proulx, M.Proulx@bath.ac.uk, University of Bath, Claverton Down, Bath, Somerset, BA2 7AY.

XXXX-XXXX/2019/9-ART1 $15.00
https://doi.org/10.1145/3352763

CCS Concepts: • **Computing methodologies → Computer vision**; **Perception**; • **Human-centered computing →** *Virtual reality*; *Empirical studies in HCI*.

Additional Key Words and Phrases: saliency, visual attention, virtual environments

ACM Reference Format:
Jacob Hadnett-Hunter, George Nicolaou, Eamonn O'Neill, and Michael Proulx. 2019. The effect of task on visual attention in interactive virtual environments. *ACM Transactions on Applied Perception* 1, 1, Article 1 (September 2019), 18 pages. https://doi.org/10.1145/3352763

1 INTRODUCTION

With the maturation of virtual environment (VE) design for gaming, simulation and the visual arts, there has been an increase in research on designing systems that predict user state and adapt to it in order to allow dynamic and immersive VE interaction [1, 39]. VEs remain mostly visual experiences, and scene designers often share the goal of photographers and film directors to guide and direct the attention of their viewers [48]. Yet, while the allocation of human visual attention has been heavily studied in static imagery [5, 42] and video [35], it remains relatively unstudied in VEs.

Understanding and predicting the allocation of human visual attention within VEs has the potential to enable a range of important applications. Approaches such as level-of-detail reduction [36] and foveated rendering [50] can reduce computational load. These approaches in combination with foveated ray-tracing [56] provide the opportunity to create higher quality rendered VEs with minimal impact on performance. Further, future fixation prediction could be used for improved attentional re-direction [12] or dynamic event triggering in virtual worlds.

A promising approach to predicting and directing visual attention is to base these models on psychophysical evidence. Early models of human vision such as the model proposed by Itti *et al.* [28] and more recent machine learning methods [9] take this approach by learning from recorded fixations of real participants viewing visual stimuli. However, visual attention entails a complex interaction between bottom-up saliency and task dependent feature guidance and spatial bias [60]. Hence, it makes sense to study further how the tasks that a user may perform within a VE affect their visual attention. In doing so, we gain not just the understanding to build more predictive models of visual attention, but also models predicting user task from visual attention [3, 10, 18, 24], helping to realize the aim of adaptive, dynamic VEs.

With this aim, we designed a study to investigate the effect of user task on visual attention in interactive VEs. We used fixation data as a proxy for overt visual attention, and recorded fixations from participants completing freeview, search and navigation tasks within three different VEs. Differences between visual attention in task conditions were quantified by computing the predictive performance of a low-level saliency map [28] and its corresponding color, intensity and orientation feature conspicuity maps. Further, differences in fixations were quantified in terms of spatial bias, duration, depth, saccade amplitude, and whether or not they were directed towards the floor.

2 BACKGROUND

It is widely accepted that early stages of visual attention are feature based. Models such as Feature Integration Theory [53] process regions of a scene that are highly conspicuous with respect to a set of visual features. The visual scene is split into a set of topographical maps, each representing a distinct visual feature. A contrasting operation (e.g. center-surround [15]) can be conducted over the feature maps to generate a conspicuity map for each feature channel. The conspicuity maps are then combined into a single saliency map, which informs the response or the allocation of visual attention. The capability of visual features to guide the allocation of visual attention is often described by psychophysical visual search studies in which participants are asked to respond to stimuli that are differentiated by one or more visual features and situated among a field of distractor stimuli

[59, 60]. If participants can efficiently locate target stimuli which differ from the distractors in a single visual feature, then that is considered evidence for the feature's attentional guidance capability. The set of visual features used by the human visual system to allocate visual attention is not fully known but it is generally accepted that color, intensity and orientation are among them [28, 53]. Other features also probably guide visual attention, such as depth [41], size [46] and motion [37]. Wolfe and Horowitz provide an overview of certain, probable, possible and unlikely visual features to guide attention in the human visual system [60].

While it is standard to study low-level visual attention via target-distractor visual search experiments, it becomes difficult to extrapolate these results to more natural viewing conditions such as viewing complex imagery, video and day-to-day real life visual attention. An alternative approach to these studies is to observe human allocation of visual attention on more realistic stimuli and relate the observed data to differences in the stimuli [57, 61]. This approach relies on the assumption that the saccade-fixation response of the human visual system, in which the eyes quickly shift to a new location (saccade) and then linger (fixation), is a good overt proxy for the allocation of visual attention [42]. While it has been demonstrated that the location of covert visual attention can be spatially separate from fixation location, studies have also demonstrated that overt visual attention is a necessary requirement for a new fixation location [11]. This means that if a person has fixated on an area, they must have attended to it, even if they do not for the entire duration of the fixation. Hence, using fixation data is suitable as a proxy for the allocation of visual attention.

Early models of visual attention, such as Koch and Ullman's [32], were based on Feature Integration Theory and psychophysical evidence of feature contribution to early vision. The seminal work of the Itti *et al.* [28] saliency model utilised an implementation of the center-surround hypothesis [15] and color, intensity and orientation feature maps to generate multi-scale feature conspicuity maps based on local feature contrast. These maps were then normalized and combined to create a single saliency map of the visual field. Models such as these have been shown to positively and accurately predict the allocation of visual attention of users viewing static imagery [42], and have since been adapted to video with the inclusion of temporal features such as motion and flicker [35].

The relative saliency of a region in the visual field is not the sole factor in determining whether or not it will be fixated upon. While it has been demonstrated that scene saliency correlates with the allocation of visual attention in some situations [30, 42], it has also been observed that other factors are much more predictive. Henderson *et al.* [22] demonstrated that visual saliency plays a lesser role in the allocation of visual attention when in active search compared to a freeview task. Henderson and Hayes further demonstrated that scene semantic properties were a better predictor in such cases [23]. Wolfe [58] updated his Guided Search theory, a successor to Feature Integration Theory, to include the influence of top-down factors on the allocation of visual attention. Guided Search postulates that visual attention is the result of bottom-up scene saliency in combination with top-down biases and modulations. It is commonly held that top-down influence works in two ways: spatial biasing [29] and feature guidance [60]. Spatial biasing occurs when a viewer has some knowledge of the scene, that is, where they might typically find something if they have to search for it. Top-down feature selection means that areas of the visual field that have similar visual features to those of a search target will stand out; for example, green items stand out when one is asked to look for something described as green. Saliency, however, has still been shown to predict attention even in the presence of a strong task, such as search for a conjunction of features [45]. These theories and findings indicate that visual attention is the result of a complex interaction between bottom-up and top-down factors, producing different results in different contexts.

The task that a user has is an obvious way to impart top-down influence on visual attention. The task of a viewer has been demonstrated to change their visual attention within a scene. Yarbus [4, 62] showed that the gaze patterns of participants viewing a painting differed depending of whether they were freeviewing it, or asked to infer or remember facts about the scene that was depicted. Further, Land and Hayhoe qualitatively showed that task relevant information guided eye movements while making a sandwich or boiling the kettle [21, 34]. Gramann *et al.* demonstrated a tendency for participants navigating in a virtual tunnel to gaze at the central

point of the tunnel when moving forward, and its outer edges when turning. Task has also been shown to alter eye movements quantitatively. Low-level features present at fixation locations vary considerably with task [51]. Mills *et al.* [38] showed that task affected visual attention spatially, with saccade amplitudes being significantly smaller in a freeview task than in memorization and search tasks. Temporally, task has been shown to change fixation duration on particular objects [7] and in general [38], with participants fixating longer when freeviewing than when searching.

The majority of research on human visual attention has been limited to the study of attention with static imagery, including how task affects visual attention. With the exception of the work of Land, in which attention was studied in the real world, the majority of visual attention research has relied on participants viewing static images. Datasets such as MIT300 [5] and CAT2000 [30] are predominantly used to train and test new models predicting visual attention, including state-of-the-art deep learning approaches such as MLNet [9] and DeepFix [33]. There has been some research into visual attention in VEs. Peters and Itti [43] computed the predictive performance of a low-level saliency map, feature maps and several other heuristics from a dataset of participants playing GameCube games. While they attempted to split the dataset and analysis in terms of 'racing' and 'exploration' games, a systematic study of task within the same environment was not conducted. Hillaire *et al.* studied gaze behavior of participants turning in VEs, identifying similarities with observed real-life behavior [25]. Lee *et al.* [36] and Hillaire *et al.* [26] implemented visual attention prediction models within 3D graphics engines, producing a prediction of where people will look in real-time. However, while these models incorporate a mechanism for biasing prediction towards 'task-relevant' objects, they do not consider the effect of task on feature selection or spatial and temporal factors of visual attention allocation.

While there has been some success in adapting previous models to new media [27], it is not guaranteed to work in other scenarios. Tatler *et al.* [52] argue that this approach of modifying the image saliency framework is problematic and that models developed from static imagery are unlikely to generalize to other media or interaction. It is therefore important to study how visual attention varies in VEs and under different tasks. This paper investigates visual attention in interactive VEs, under a set of different tasks and in different scenes.

3 METHODOLOGY

3.1 Overview

The study had a within-participants single factor design with task (Freeview, Search or Navigation) as the primary independent variable and recorded fixations as the dependent variable. All 3 tasks were performed in each scene, both to control for the effect of scene and to add variance in the VE that should make the findings more generalizable. Participants completed five trials of each task in three different VEs (see Figure 1 and 3), resulting in a total of 45 trials. The five different trials per scene used a different start location, along with a different search item and end location for the search and navigation tasks respectively. This meant that exposure to scenes was not limited to only one view point, helping to ensure any results are not biased by a particular view. Trials were randomized across scene, task and starting location.

For each trial participants had to complete one of the three tasks. Participants were informed of their next task via a black screen with text shown between trials. In the case of the Freeview task, participants were shown "Freeview" and a top-down perspective map of the environment in which they were to be placed. This map contained a red dot indicating the location in which they were to spawn. Participants were told prior to the study that in the Freeview task they were to view and observe the environment at their own free will. Each Freeview task took 30 seconds. For the Search task, participants were shown the text "Search" and the same top-down map of the environment with their future spawn location. They were also shown a small image of their search target [55]. We showed participants an image of the target rather than describing it textually in order to prevent biasing their attention by the language we used to describe it. For example, had we asked participants to

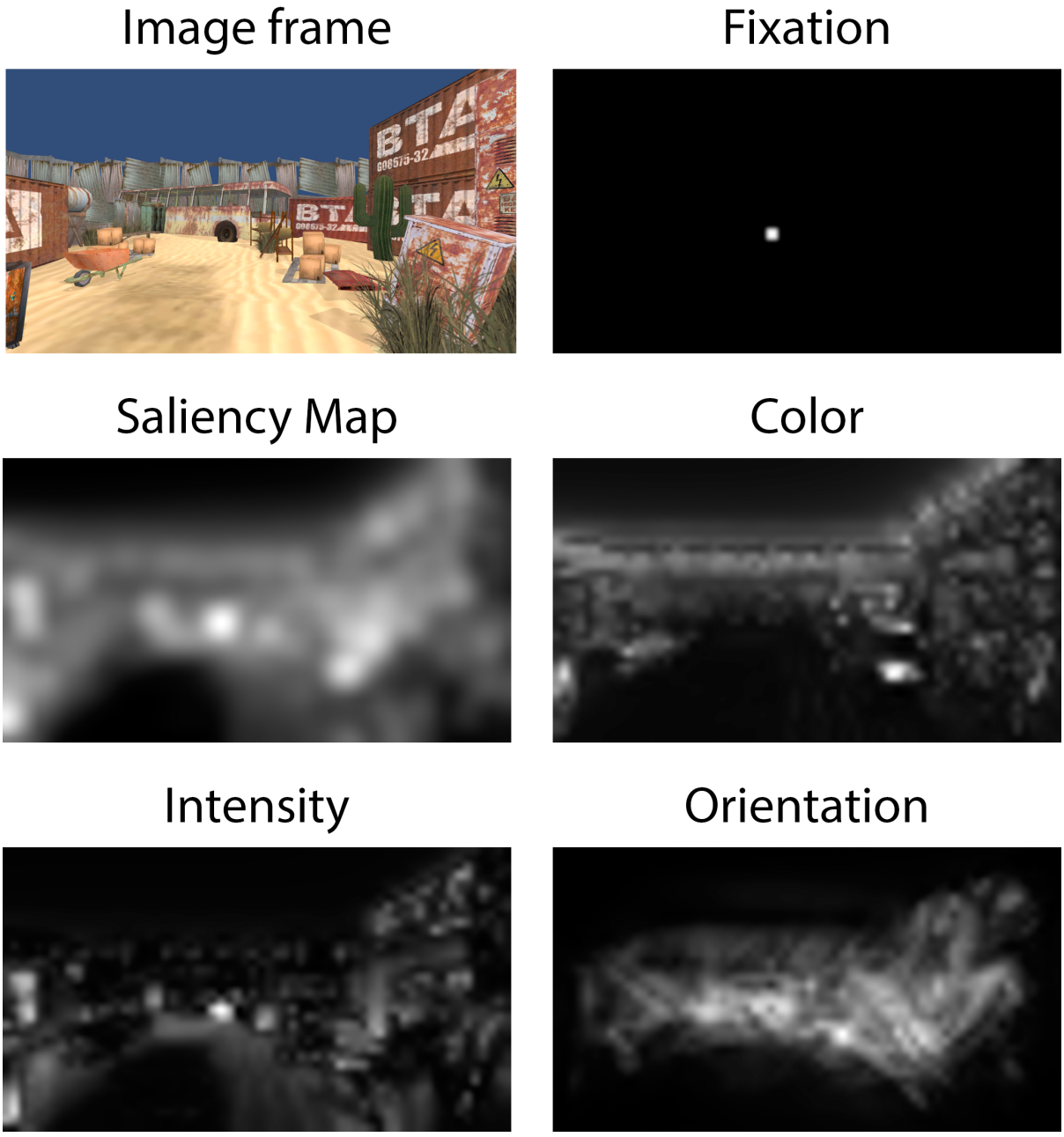

Fig. 2. An example frame from the the Junkyard VE along with a participant's fixation location, the generated saliency map and color, intensity and orientation conspicuity maps.

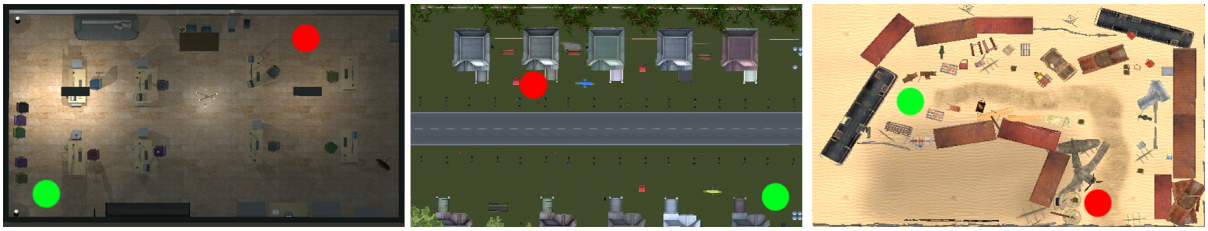

Fig. 3. Top-down orthographic perspectives of the three scenes. On the left is an indoor office space. In the middle, a suburban street. On the right is a desert junkyard. The green and red dots indicate example start and end points respectively for the Navigation condition

search for "the red sock" and "the green basketball" we might have unwittingly biased attention towards color features [49]. Participants were informed that in the search task they should actively seek out the object within the environment and move towards it, ending the trial. Finally, for the Navigation task participants were shown the text "Navigation" and the map containing the red dot for their spawn location, but also a green dot for their target location. Participants were informed that during Navigation tasks they should make their way from their spawn location to the endpoint as quickly as possible.

The role of the top-down view map before each trial was two-fold. First, it facilitated a natural navigation task for the Navigation condition. Participants had to move from a start point to an end point indicated on their map view of the environment. This aimed to replicate how people navigate unknown environments in the real world, often using their phones and map apps, or GPS systems in cars. Secondly, the top-down map was used as a control in all conditions to match the stimuli and information provided to participants, and also to ensure that participants did not need to spend initial time orienting themselves within the environment. For example, in the Search task, participants knew where they were and thus could immediately focus on searching for their target object.

3.2 Participants

There were 19 participants (12 male, 7 female) with a mean age of 21.9 (SD=1.2). Participants were convenience sampled from the undergraduate population at the University of [REDACTED]. All participants were naive to the purpose of the study, and none had seen the scenes before. All participants had normal or corrected to normal vision. All participants provided written, informed consent for public distribution of collected data.

3.3 Apparatus

All VEs were presented to the participants using the same hardware and environmental setting. The study took place in a quiet, secluded room with no interference and no visual or auditory distraction. Participants sat on a standard office swivel chair with adjustable height both for comfort and to ensure that their eyes were in the centre of the screen. Participants viewed the VEs on a 23" (20.05 inches x 11.28 inches) monitor with a 60Hz refresh rate and 5ms typical response time. A chin rest was used to ensure the participant's head was central to the screen and their eyes remained 60cm from the screen. This setup resulted in a horizontal viewing angle of 46.0° and a vertical viewing angle of 26.9°. Participants interacted with the VEs via mouse and keyboard, placed in front of them on the same desk as the monitor. A Tobii TX300 eye tracker was integrated with the monitor and used to record participants' gaze position at 300Hz.

3.4 Materials

In order to ensure that any task differences were not unique to the scene in which they were performed, we developed three different VEs for the experiment to take place in. Virtual Junkyard, Street and Office environments were created using a wide variety of 3D assets that were consistent in style and appropriate to the scene in which they were placed. The scenes were rendered in real-time using the Unity3D game engine. Participants viewed the scene from a first person perspective, controlling the yaw and pitch of their camera with the mouse, and their camera position using the W, A, S and D keys to move forwards, left, backwards and right respectively. This is a typical interaction mechanism used within first person VEs, common in PC games and simulation experiences.

The scene assets were selected such that the scenes contained a variety of visual stimuli. Assets varied substantially in color, texture, intensity, size and orientation of placement. This variety in visual stimuli serves an important purpose. First, the wide variety of assets mimics what one would see in real scenes, meaning that visual attention is less likely to be biased by the limited selection of objects in a typically bare experimental scene, and the results should therefore be more generalizable to visual attention in interactive VEs such as games and training environments. Secondly, by including objects that vary in many different visual features, we reduce the risk of biasing attention to any one feature. Further, scene assets were consistent in style and appropriate to the scene in which they were placed. This was to avoid attention being biased towards certain objects simply because they were unusual for the context.

All three environments were also relatively cluttered, with a lot of small and larger objects scattered throughout the scene. The Office scene, for example, featured several desks in the middle of the room, all covered in office equipment and books. The Junkyard scene contained scattered industrial equipment and disused household appliances on the floor. In the freeview task, we wanted participants to have a variety of competing stimuli for them to look at, keeping them engaged and exploring. For the search task, we also wanted a variety of competing stimuli to make the search harder, ensuring that participants were actively searching during the majority of the task. If the search was too easy, then participants would easily find the object and simply navigate towards it, effectively making the search task another version of the navigation task. Finally, for the navigation task we wanted participants to need to actively navigate and avoid obstacles to prevent the task from being too easy.

3.5 **Procedure**

Participants were briefed and signed a consent form prior to taking part in the study. To get used to the keyboard and mouse controls, participants were given up to 5 minutes to explore a simple scene which shared no assets with the test environments. Participants were then told what they had to do for each of three tasks, and that the trials would transition with them being informed of their task and any relevant information prior to the scene loading. When participants were ready to start the study they placed their chin on the chin rest. A 5 point calibration process was used to calibrate the eye tracker, the experimental software was launched and the participant was given control of the mouse and keyboard.

3.6 **Data Analysis**

The Tobii TX300 eye tracker recorded participant gaze points at 300Hz. In order to classify these gaze points into fixations, we used Tobii Studio's I-VT fixation classifier [40], with adjacent gaze recordings with a velocity under 30 deg/s being classified as fixations, and adjacent fixations within 0.5 degrees and 75ms being combined. Fixations of duration less than 60ms were discarded. This process resulted in 56,193 fixations across all participants and trials, defined in terms of X and Y screen pixel coordinates and fixation duration.

During the study, participant camera position and angles were saved every frame, allowing us to re-render everything they observed without recording the screen during the experiment. We matched camera positions and rotations to fixation data and saved out rendered image frames corresponding to each fixation. For each

image frame we computed a saliency map using an implementation of the saliency model proposed by Itti *et al.* [28] and implemented by Harel *et al.* [19, 20] in their saliency toolbox. This saliency model computes a saliency map via combination of three feature conspicuity channels: color, intensity and orientation. We generated these corresponding feature conspicuity maps for each fixation frame prior to them being combined into the final saliency map. Figure 2 shows example maps. In this model, all feature conspicuities are computed using an implementation of the center-surround hypothesis in which the input image is repeatedly sub-sampled and blurred to several levels of input from fine to coarse detail. Feature maps are then generated for each feature and each level, and the difference operation between coarse and fine levels ultimately produces a single map for each feature channel representing multi-scale conspicuity. In this model, color conspicuity is produced from two sub-features, red-green color opponency and blue-yellow color opponency. Intensity feature maps are generated by taking the average of red, green and blue color channels of the input images. Orientation conspicuity is generated from 4 different sub-orientation features, computed via convolving Gabor filters representing 0, 45, 90, and 135 degrees with the input images. By outputting feature conspicuity maps separately we are able to quantify the predictive performance of them separately, and thus potentially identify any difference in feature reliance between tasks.

To quantify the performance of saliency and feature maps, and thus quantify the reliance of participants on saliency and feature information to allocate their visual attention, we produce Normalized Scanpath Saliency (NSS) [6, 44] scores for each map. NSS measures the activation assigned to a fixation i on a saliency (or feature) map S that has been normalized to have zero mean unit standard deviation:

$$NSS(i) = \frac{S(i) - \mu(S)}{\sigma(S)}$$

We compute the NSS score of all 56,193 fixations and their corresponding saliency maps and feature conspicuity maps. In practice, we group these NSS scores by participant, task and scene, averaging fixation within a participant to produce a mean NSS (MNSS) score for each participant in each potential condition. A score of zero indicates that there is no correlation between the map and fixation, a positive score indicates positive correlation (good performance) and negative score indicates negative correlation (poor performance).

To further analyse the allocation of visual attention depending on task, we also studied spatial and temporal aspects of the fixations themselves. Fixation duration was provided by the eyetracker software and we compared fixation durations between tasks. When outputting image frames from the Unity Game Engine, we also raycasted fixations from the camera position into the scene and recorded the distance from camera to the point of collision. This allowed us to study how depth of fixation varies with task. Further, we outputted the asset name of the object with which the ray collided, identifying which objects participants were fixating on at any given time. Consistent semantic labels between scenes were hard to identify, but one labelling that was consistent and potentially useful for understanding navigation was determining if a fixation was a 'floor' or 'not floor' fixation. This labelling allowed us to determine the percentage of fixations spent looking at the floor. Finally, for each fixation we also computed a measure of 'saccade amplitude' [2] by calculating the euclidean distance between each fixation and the previous one. Saccade amplitude has been shown to vary with task [7], and so we computed this measure to study its variability in tasks within VEs.

To analyze differences between tasks with respect to these measures of visual attention, we ran a set of analysis of variance (ANOVA) tests. We grouped all fixations by task and averaged measures by participant. One-way ANOVAs indicated the overall effect of task on these measures, and we ran further post hoc t-tests with Bonferroni corrected p-values for comparison of specific tasks.

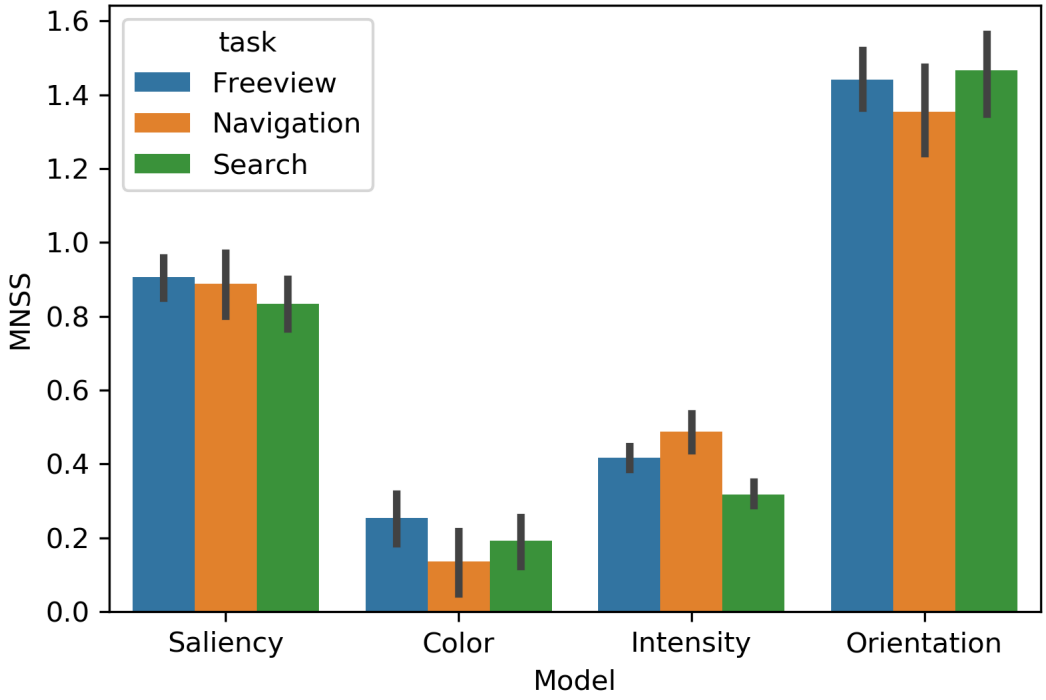

Fig. 4. Mean NSS results for Saliency, Color, Intensity and Orientation maps. Fixation NSS scores for each map were grouped by task and participants, with the height of the bars representing the average participant's MNSS score. Error bars represent the 95% confidence interval.

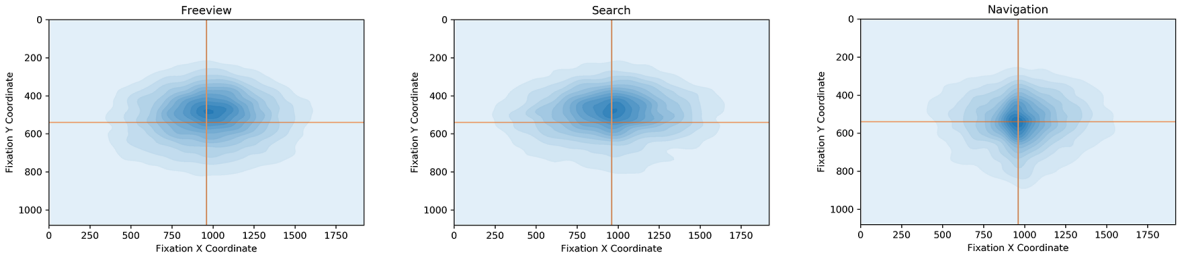

Fig. 5. Kernel Density Estimate graphs of fixation X and Y coordinates in the Freeview, Search and Navigation task respectively. The intersection of the lines represents the center point of the visual field (screen).

4 RESULTS

4.1 Saliency and Feature Reliance

Figure 4 shows the MNSS for Saliency, Color, Intensity and Orientation maps. Saliency MNSS scores for Freeview, Search and Navigation tasks were 0.9059, 0.8340 and 0.8878 respectively. There were no statistically significant differences between task means as determined by one-way ANOVA ($F(2,54) = 0.974$, $p = 0.3842$). Post hoc analyses using a Bonferroni correction found that Freeview produced significantly higher Saliency MNSS scores than the Search condition ($t = 3.852$, $p = 0.0037$), albeit with a small effect size. The Navigation task did not produce significantly different results from Freeview ($t = -1.657$, $p = 0.3444$), or Search ($t = 0.554$, $p = 1$).

For the Color feature, there were no statistically significant differences between task means as determined by one-way ANOVA ($F(2,54) = 2.455$, $p = 0.0953$), however, post hoc analysis showed that Freeview produced significantly higher MNSS scores than Search ($t = 4.549$, $p = 0.0007$) and Navigation ($t = 3.754$, $p = 0.004$). Search and Navigation were not significantly different ($t = 1.753$, $p = 0.2897$).

For the Intensity feature, there was a statistically significant difference between task means as determined by one-way ANOVA ($F(2,54) = 18.155$, $p < 0.0001$). Navigation task produced the highest MNSS scores, significantly higher than Search ($t = -6.204$, $p < 0.0001$) but not Freeview ($t = 2.251$, $p = 0.062$) . The Search condition produced the second highest Intensity MNSS scores, significantly higher than Freeview ($t = 6.420$, $p < 0.0001$).

Finally, for the Orientation feature, there was no statistically significant difference between task means as determined by one-way ANOVA ($F(2,54) = 1.213$, $p = 0.305$). Post hoc analysis found that the Navigation condition produced significantly smaller MNSS scores than Search ($t = 2.892$, $p = 0.0292$) but not Freeview ($t = 2.421$, $p = 0.079$). Freeview and Search conditions did not produce significantly different Orientation results ($t = -0.920$, $p = 1$).

4.2 Spatial and temporal bias of fixations

Figure 5 shows Kernel Density Estimate plots of fixations split between Freeview, Search and Navigation tasks with screen center lines of X = 960 and Y = 540 plotted for reference. Areas of darker blue represent higher density areas of fixation. Visual inspection identifies a strong center bias for fixations in all conditions. The KDE plots suggest a slight bias towards the upper visual field for the Freeview and Search tasks, while the Navigation task produces more central results. This is confirmed by participant mean fixation Y coordinates for the Navigation task (543) being significantly larger than Freeview (516) ($t = -4.106$, $p = 0.0019$) and Search (501) ($t = 5.063$, $p = 0.0002$). Regarding the X coordinates, Freeview produced a mean value of 972, Search a value of 946, and Navigation a value of 970. There was no significant difference between tasks as determined by one-way ANOVA ($F(2,54) = 2.136$, $p = 0.1279$). Further, fixations for the Navigation task seem have a smaller spread, and thus are more consistent in their central bias. This is validated by computing the determinant of the co-variance matrices for fixation X and Y coordinates for each participant in each condition and averaging across participants as a measure for dispersion. Search produced the highest co-variance matrix determinant (1485932204), Freeview the second highest (1307289078), and Navigation the least (1087900569). Navigation dispersion was significantly lower than Search ($t = 3.063$, $p = 0.0201$), but not Freeview ($t = 1.924$, $p = 0.2108$). Freeview and Search dispersion were however significantly different ($t = 2.737$, $p = 0.0406$).

Further regarding the spatial allocation of fixations, we labelled objects that fixations landed on as either ‘floor’ or ‘not floor’. We observed a significant difference between tasks in percentage of fixations allocated towards the floor as determined by one-way ANOVA ($F(2,54) = 10.706$, $p = 0.0001$). Freeview produced the least mean percentage of fixations towards the floor with 26.3%, followed by Search with 34.3%, and Navigation with 42.2%. Percentages were significantly different between all conditions ($p < 0.0001$).

We also computed the mean saccade amplitude, measured in terms of the euclidean distance in pixels between each fixation and the previous one, for each task. A one-way ANOVA ($F(2,54) = 4.589$, $p = 0.0144$) identified a

significant effect of task on saccade amplitude. The mean saccade amplitude was lowest for the Navigation task at 189 pixels, which was significantly smaller that the mean saccade amplitude of Freeview with 203 pixels ($t = 2.661$, $p = 0.0476$). Search produced the highest mean saccade amplitude of 219 pixels, significantly higher than Navigation ($t = -3.900$, $p = 0.0031$) and Freeview ($t = -3.556$, $p = 0.0067$).

The depth of each fixation was also calculated. A one-way ANOVA ($F(2,54) = 48.175$, $p < 0.0001$) identified a significant effect of task on fixation depth. The mean distance of fixation into the scene was smallest for the Freeview task (9.67m), second for the Search task (10.61m) and highest for the Navigation task (15.54m). Navigation mean fixation depth was significantly higher than both Freeview and Search ($p < 0.0001$). While Search and Freeview fixation depths were close together, Search depths were significantly deeper into the scene ($t = 3.557$, $p = 0.0067$).

We also compared the fixation duration between tasks. A one-way ANOVA ($F(2,54) = 4.298$, $p = 0.0185$) showed a significant effect of task on fixation duration. Post hoc analysis identified that the mean duration of fixations in Freeview (226ms) and Search (219ms) were not significantly different ($t = 1.379$, $p = 0.5511$). Navigation, however, produced the highest mean fixation duration of 269ms, significantly higher than both Freeview and Search ($p < 0.0001$).

4.3 Interactions with scene

A two-way ANOVA was conducted to study the effect of task and scene on fixation duration. The significant effect of task on fixation duration was confirmed when fixations were grouped by task and scene ($F(2,168) = 10.595$, $p < 0.0001$). The main effect for scene, however, was not significant ($F(2,168) = 1.365$, $p = 0.2582$). There was no significant interaction effect between task and scene ($F(4,168) = 0.181$, $p = 0.9481$).

A two-way ANOVA produced similar results for the effect of task and scene on saccade amplitude. There was a significant main effect of task ($F(2,168) = 10.873$, $p < 0.0001$) but not of scene ($F(2,168) = 0.605$, $p = 0.5474$). There was no significant interaction effect between task and scene ($F(4,168) = 1.765$, $p = 0.1384$).

All effects on depth of fixation were significant. The main effect of task produced an F ratio of $F(2,168) = 58.093476$, $p < 0.0001$. The main effect of scene produced an F ratio of $F(2,168) = 417.149$, $p < 0.0001$. The interaction of task and scene produced an F ratio of $F(4,168) = 332.578$, $p < 0.001$. We ran a further two-way ANOVA in which each fixation depth was normalized by the maximum observed fixation distance within the scene in which it occurred. This normalisation produced the same highly significant main effects ($p < 0.0001$), however, did not find a significant interaction effect ($F(4,168) = 2.419$, $p = 0.051$).

Finally, regarding fixations being classified as fixating on the floor or not, all effects were significant. Task produced a significant main effect ($F(2,168) = 25.454$, $p < 0.0001$). Scene also produced a significant main effect ($F(2,168) = 12.335$, $p < 0.0001$). There was a significant interaction effect between task and scene ($F(4,168) = 3.120$, $p = 0.0166$).

4.4 Effect of participant gaming experience

To identify whether the input modality used in this study produced any effect on participant viewing behaviour, we analysed the variance of attention measures with respect to self-reported gaming experience. The mean reported gaming experience was 3.05 ($SD = 1.12$). In general, we observed no significant effect of gaming experience on the attention metrics used above. A one-way ANOVA identified no significant effect of self-reported gaming experience on Saliency MNSS ($F(4, 14) = 2.964$, $p = 0.0575$), fixation duration ($F(4, 14) = 1.042$, $p = 0.4204$), saccade amplitude ($F(4, 14) = 1.428$, $p = 0.2761$) and floor fixations ($F(4, 14) = 2.249$, $p = 0.1158$).

5 DISCUSSION

5.1 Saliency and feature conspiciuty in the allocation of visual attention

In order to investigate the reliance of human visual attention allocation on saliency and feature conspicuity information depending on task, we computed the MNSS score for Freeview, Search and Navigation tasks. All tasks produced positive MNSS scores, indicating a positive reliance on saliency information to allocate visual attention in all tasks. Perhaps unexpectedly, however, there was little to no significant difference of saliency MNSS scores between tasks. While Freeview produced the highest MNSS score, Navigation second and Search last, the only significant difference was between Search and Freeview. Higher MNSS scores for Freeview than for Search makes intuitive sense. The absence of an explicit task in the Freeview condition could reduce top-down influence on the bottom-up processing of visual information, thus increasing reliance on saliency. Search producing the lowest MNSS scores could also have been expected. As participants were required to look for a specific object that they had been shown, it could be expected that Search would produce the largest top-down influence on attention as participants were seeking specific image features that matched their search target. The lack of a statistically observable difference between Navigation and Freeview could be explained by the fact that navigation induces top-down influence in a less specific way than Search, requiring participants to follow a navigable path and avoid obstacles. In this case, directing attention to salient changes in peripheral vision could be used as a mechanism for staying on path.

A possible explanation for the small differences in saliency MNSS scores between tasks is that task may affect visual attention by biasing it towards certain feature conspicuities. It is possible that the overall saliency reliance as computed by the Itti *et al.* [28] saliency map MNSS scores could remain relatively constant; that is, if certain feature conspicuity channels predict attention more, then other channels predict it less. The feature conspicuity results, as seen in Figure 4, provide an example of this. Navigation and Freeview saliency MNSS scores are not significantly different, yet Navigation produced significantly smaller color and orientation MNSS scores but significantly higher intensity MNSS scores. As the saliency map is a normalized combination of normalized color, intensity and orientation conspicuity maps, this feature reliance variance between tasks should contribute to the small difference in saliency reliance.

The variance in feature conspicuity MNSS scores further indicates that different tasks produce different relative reliance on feature channels for the allocation of visual attention in VEs. This result is interesting for several reasons. First, it suggests a possible adaptation of the Itti *et al.* [28] model in which feature channel contribution to the final saliency map is weighted by the expected influence of each feature, similar to the approach taken by Zhao and Koch [63]. While altering the model in this way would make it no longer a pure model of saliency, it would improve its ability to predict the allocation of human visual attention.

Some of the results make intuitive sense, while others are less obviously explainable. The relatively smaller reliance of the navigation task on orientation conspicuity to allocate visual attention initially seems unintuitive. It could be reasoned that orientation reliance should be higher than in other tasks, due to significant orientation changes correlating with object features such as edges and ledges. These are things that a navigating human definitely needs to be aware of in order to avoid. However, it has been suggested that obstacle fixation is not required for navigation and obstacle avoidance, with people being capable of effectively navigating while mostly keeping potential obstacles within their peripheral vision [13]. Further, we observed that the navigation task seems to rely less on color conspicuity than other tasks. Lack of attention to color conspicuity is understandable due to all three scenes containing a variety of assets with different, but natural, coloring. Color conspicuity therefore would correlate less with pathways and obstacles, reducing its usefulness for successful navigation. The navigation task also relied relatively more on intensity conspicuity than the other tasks. To the best of our knowledge, this result has not been previously published. Why intensity information predicts attention in a

navigation task better than in other tasks remains unknown but it is an interesting result and potentially useful for the applications of understanding feature reliance variability discussed above.

5.2 Task produces a spatial and temporal bias on fixation allocation

Our results demonstrated a significant impact of task on spatial and temporal aspects of fixation. In short, navigating participants fixated more centrally, for longer, and deeper into the VEs. Shifts between their fixations were smaller, and they spent more time looking at the floor than when completing other tasks. In general, these measures show smaller effects between the Freeview and Search tasks or, in the case of fixation duration, no significant difference. These results demonstrate that, with regards to the spatial and temporal aspects, navigation tasks produce a more targeted and focused allocation of attention. This result is consonant with previous work on visual attention which demonstrated an object bias depending on whether participants were navigating towards or avoiding certain objects [47] and a bias of attention towards the central path [16].

Our analysis of the spatial distribution of fixations identified several key differences between tasks. Navigation fixations were more centrally biased: the mean fixation location was closer to the center of the visual field and fixations were less dispersed than in other tasks. The increased center bias for fixation in the navigation condition may be influenced by the interaction mechanism. The first-person mouse and keyboard interaction meant that, when moving forward, the camera would move parallel to the floor plane in the direction of the camera's yaw vector, so participants moved in the direction that their 'virtual head' was pointing. As participants were navigating, likely focusing on goal locations, they may have been relying on head movements and centering their goal locating within the center of the screen instead of fixating towards the edges of their visual field. The higher dispersion of attention for the Freeview and Search tasks also makes sense. Freeviewing participants had no extrinsic goal and would conceivably use more of the visual field to explore and find interesting things to observe. Nonetheless, center bias has been observed in freeviewing visual attention datatsets [54], and so we would still expect to observe it here. For the Search task, participants received no prior information about the places that the object might reside. It is possible that participants explored more of their visual field in order to search areas that contained features matching those of their target object.

Complementing the center bias results are our findings on how saccade amplitude varies between tasks. Navigating participants shift their fixation points the least, followed by freeviewing participants and searching participants. These results correspond to the center biases in each task. Navigating participants are more centrally fixated, and so it would be expected that the distance from one fixation to the next would be smaller. Searching participants were exploring their visual field the most, thus leading to bigger shifts of attention. Further, navigating participants fixated significantly deeper into the scenes and the duration of these fixations was significantly longer. These results in tandem demonstrate a very different viewing strategy used by participants during Navigation than during other tasks, to fixate towards a far central landmark. Similar viewing strategies have been observed previously. For example, participants navigating a virtual tunnel gazed mostly towards the centroid of the tunnel [16]. Coupled with the longer and deeper fixations, the observation that navigating participants looked more at the floor than in other tasks suggests that they may have been fixating on their goal location, or distant intermediate waypoints, in order to navigate. Stated simply, our results suggest that navigating participants look where they are going. This is not unexpected, however, what is important is that this viewing strategy for navigation is observable and distinguishable from the viewing strategies of other tasks. The implication is that, for the development of visual attention models, biasing predictions towards distant floor regions in the center of the visual field may improve predictive performance when the user is navigating.

5.3 Generalizability of attention measures

Our results have demonstrated effects of task on several measures of human visual attention. Our initial analysis grouped fixations by task, disregarding the effect of scene. By combining fixations from different scenes, we introduced variance into our dataset. The fact that we still see significant results indicates that there is a consistent and strong effect of task on visual attention across scenes, supporting our belief that these results may be generalized to dynamic VEs.

Fixation duration and saccade amplitude generalized extremely well, demonstrating only a main effect of task with no significant main effect of scene and no interaction effect. This suggests that variance in these measures is dependent only on what the user is doing, and not on where they are doing it. Both of these measures have been shown to vary with task when viewing static imagery [7, 38], however, to the best of our knowledge the effect of task on these measures has not previously been demonstrated for tasks performed within dynamic three dimensional environments. Further, identifying that these measures are affected by task independently of where that task takes place is extremely useful. These measures can thus be considered good indicators of what task a person is performing, and therefore good features for models attempting to identify user task from visual attention [17].

Some measures did show a main effect for scene and an interaction effect. Depth of fixation showed effects of task and scene and an interaction effect between the two. A reason for this is that different scenes had different levels of depth. The road scene extended much further and was more spread out than the office scene. If participants were looking at objects in the scene, or their goal locations, then deeper fixations would occur in a larger scene. Normalizing the fixations by the maximum observed fixation within each scene confirmed this. The main effects were still present, as scene size will still affect fixation depth, but the interaction effect was removed. This indicates that depth of fixation is dependent on the user's task and where they are performing that task, but that the relative effect of task does not change with scene provided you control for scene size. Similarly, we identified a main effect of scene on the amount of fixation directed towards the floor. Again, this is explained by some scenes having more floor space and less clutter than others. Nonetheless, the main effect of task was still present, again indicating this feature's usefulness in terms of predicting attention (i.e. biasing predictions towards floor regions when navigating) or identifying user task from visual attention data.

5.4 Limitations and further work

There are a couple of potential limitations to this study to be addressed in future work. In our experimental design we ensured that all participants viewed the scenes while using a chin rest, restricting natural head movements of participants in our study. There were several benefits to this approach, including increased quality of eye tracking data and ensuring that all participants viewed the VEs with the same visual angles. However, several works have demonstrated head-gaze movement interactions under natural viewing conditions [8, 31]. The restriction of head movements may have affected human visual behavior in this study, for example increasing center bias of fixations. The reported findings however still hold validity. Firstly, participants used the chin rest in all conditions, meaning the observed impact of task on visual attention is still meaningful. Secondly, the setup we utilized for this study meant that all screen regions could be comfortably fixated on without the need for a head movement. The screen covered a horizontal viewing angle of 46.0° and a vertical viewing angle of 26.9°, well bellow the 80° human limit for saccade without head movement [14].

Our choice of input modality also impacts the interpretation of these result. Mouse and keyboard controls are a common input modality for VEs yet nonetheless impose an unnatural layer of interaction that is not present during natural bodily and head movement based interaction with real world environments. Our choice of input modality was dictated by the size of environments and task set we expected participants to carry out. A mouse and keyboard offers finer and more fluid control of camera angle than joysticks on a controller. A head-mounted

display VR setup would have allowed for more natural head and body movement, but it would have limited the large navigable size of the environments needed for the navigation and search tasks. We hypothesize that many gaze behaviors carry over from natural viewing to VE viewing regardless of input modality. By having our participants use a chin rest we enforced head movement via mouse movement, meaning that the movements altering visual field afforded to the participants, and the degrees-of-freedom of those movements, remained close to real life.

Regardless of the natural validity of mouse and keyboard controls there are also implications of input modality on visual behavior specifically within VEs. It is possible that those with more experience with the input modality would have more control and thus exhibit different visual behavior. However, all participants received a training period ensuring they were comfortable with the controls and has a basic level of competency. Further, we identified no significant effect of previous gaming experience on visual attention measures, indicating that participants who were more likely to have large amounts of experience with the input modality did not bias our results. It is still possible, however, that the input modality biased visual attention measures on the whole. For example, perhaps the central bias in our results is larger than in natural viewing conditions. If this were the case it could be due to participants being more likely to reorient their visual field by making a virtual head movement (i.e. mouse movement) than by fixating towards the edges of the visual field. Further work is required to identify the impact of input modality on both visual attention allocation and gaze-head movement interaction. We propose a future study in which visual display medium and input modality are decoupled, for example by having participants view VEs on screen with mouse controls, in a HMD with head movement controls, and in a HMD with mouse controls.

Further work is also needed in studying visual attention with respect the myriad of other visual features that can draw attention. For example, within this study motion features were not considered as there was no movement present within the scenes themselves. Any motion present was only an artifact of perspective change, and thus akin to optic flow, which has been identified as probably not being a guiding attribute of human visual attention [60]. Nonetheless, studying visual attention is scenes with external motion present is an obvious progression. Other features known to guide attention were also not included in this study. However, the visual representation of conspicuity with respect to these feature are not always as obvious as colour, intensity and orientation. Substantial further research is required to identify how best to compute and represent the full set of visual features that may affect the allocation of human visual attention.

Finally, we conducted this study over three scenes in order to increase the generalizability of our results to visual attention in virtual environments in general. We did however observe significant differences in some attention measures with respect to scene. In order to keep this work focused on task we did not study the effect of scene too deeply. Future work could focus on identifying how scene changes affect human visual attention. This would most likely require substantial new experimentation in order to de-confound semantic changes from visual feature changes between scenes.

5.5 Conclusions

In this paper we have presented a study of human visual attention in VEs and how it varies with user task. To the best of our knowledge, we have presented the first results indicating that three tasks, Freeview, Search and Navigation, produce a significant effect on participant viewing behavior across different VEs. The effect of task is particularly apparent in several measures of human visual attention when participants are navigating, indicating their usefulness for predicting user's visual attention, with potential for dynamically adapting games and simulations. Further, our results contribute to the relatively small but growing literature on human visual attention in VEs. Current state-of-the-art models of visual attention with static imagery [9] and VEs [26] do not take account of the user's task. Our findings that task significantly affects the allocation of visual attention

suggest that more consideration should be given to task in order to build better, more predictive, models of visual attention in dynamic VEs.

ACKNOWLEDGMENTS

Eamonn O'Neill's research is partly funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1.i

REFERENCES

[1] Sander Bakkes, Chek Tien Tan, and Yusuf Pisan. 2012. Personalised gaming: a motivation and overview of literature. In *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System*. ACM, 4.
[2] Robert W Baloh, Andrew W Sills, Warren E Kumley, and Vicente Honrubia. 1975. Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology* 25, 11 (1975), 1065–1065.
[3] Jonathan FG Boisvert and Neil DB Bruce. 2016. Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features. *Neurocomputing* 207 (2016), 653–668.
[4] Ali Borji and Laurent Itti. 2014. Defending Yarbus: Eye movements reveal observers' task. *Journal of vision* 14, 3 (2014), 29–29.
[5] Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. 2015. MIT saliency benchmark.
[6] Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. 2018. What do different evaluation metrics tell us about saliency models? *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018).
[7] Monica S Castelhano, Michael L Mack, and John M Henderson. 2009. Viewing task influences eye movement control during active scene perception. *Journal of Vision* 9, 3 (2009), 6–6.
[8] CJS Collins and GR Barnes. 1999. Independent control of head and gaze movements during head-free pursuit in humans. *The Journal of physiology* 515, 1 (1999), 299–314.
[9] Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. 2016. A deep multi-level network for saliency prediction. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 3488–3493.
[10] Antoine Coutrot, Janet H Hsiao, and Antoni B Chan. 2018. Scanpath modeling and classification with hidden Markov models. *Behavior research methods* 50, 1 (2018), 362–379.
[11] Heiner Deubel and Werner X Schneider. 1996. Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research* 36, 12 (1996), 1827–1837.
[12] Magy Seif El-Nasr, Athanasios Vasilakos, Chinmay Rao, and Joseph Zupko. 2009. Dynamic intelligent lighting for directing visual attention in interactive 3-D scenes. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 2 (2009), 145–153.
[13] John M Franchak and Karen E Adolph. 2010. Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research* 50, 24 (2010), 2766–2774.
[14] Edward G Freedman. 2008. Coordination of the eyes and head during visual orienting. *Experimental brain research* 190, 4 (2008), 369.
[15] Dashan Gao, Vijay Mahadevan, and Nuno Vasconcelos. 2008. On the plausibility of the discriminant center-surround hypothesis for visual saliency. *Journal of Vision* 8, 7 (2008), 13–13.
[16] Klaus Gramann, Jennifer El Sharkawy, and Heiner Deubel. 2009. Eye-movements during navigation in a virtual tunnel. *International Journal of Neuroscience* 119, 10 (2009), 1755–1778.
[17] Amin Haji-Abolhassani and James J Clark. 2014. An inverse Yarbus process: Predicting observers' task from eye movement patterns. *Vision Research* 103 (2014), 127–142.
[18] S Navid Hajimirza, Michael J Proulx, and Ebroul Izquierdo. 2012. Reading users' minds from their eyes: a method for implicit image annotation. *IEEE Transactions on Multimedia* 14, 3 (2012), 805–815.
[19] Jonathan Harel, C Koch, and P Perona. 2006. A saliency implementation in matlab. *URL: http://www. klab. caltech. edu/harel/share/gbvs. php* (2006).
[20] Jonathan Harel, Christof Koch, and Pietro Perona. 2007. Graph-based visual saliency. In *Advances in neural information processing systems*. 545–552.
[21] Mary Hayhoe and Dana Ballard. 2005. Eye movements in natural behavior. *Trends in cognitive sciences* 9, 4 (2005), 188–194.
[22] John M Henderson, James R Brockmole, Monica S Castelhano, and Michael Mack. 2007. Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye movements*. Elsevier, 537–III.
[23] John M Henderson and Taylor R Hayes. 2017. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour* 1, 10 (2017), 743.
[24] John M Henderson, Svetlana V Shinkareva, Jing Wang, Steven G Luke, and Jenn Olejarczyk. 2013. Predicting cognitive state from eye movements. *PLOS ONE* 8, 5 (2013), e64937.

[25] Sébastien Hillaire, Anatole Lécuyer, Gaspard Breton, and Tony Regia Corte. 2009. Gaze behavior and visual attention model when turning in virtual environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 43–50.
[26] Sebastien Hillaire, Anatole Lecuyer, Tony Regia-Corte, Rémi Cozot, Jerome Royan, and Gaspard Breton. 2012. Design and application of real-time visual attention model for the exploration of 3D virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 18, 3 (2012), 356–368.
[27] Laurent Itti. 2002. Real-time high-performance attention focusing in outdoors color video streams. In *Human Vision and Electronic Imaging VII*, Vol. 4662. International Society for Optics and Photonics, 235–244.
[28] Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259.
[29] Yuhong V Jiang, Z Sha Li, and Roger W Remington. 2015. Modulation of spatial attention by goals, statistical learning, and monetary reward. *Attention, Perception, & Psychophysics* 77, 7 (2015), 2189–2206.
[30] Tilke Judd, Frédo Durand, and Antonio Torralba. 2012. A benchmark of computational models of saliency to predict human fixations. (2012).
[31] Aarlenne Z Khan, Gunnar Blohm, Robert M McPeek, and Philippe Lefevre. 2009. Differential influence of attention on gaze and head movements. *Journal of neurophysiology* 101, 1 (2009), 198–206.
[32] Christof Koch and Shimon Ullman. 1987. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*. Springer, 115–141.
[33] Srinivas SS Kruthiventi, Kumar Ayush, and R Venkatesh Babu. 2017. Deepfix: A fully convolutional neural network for predicting human eye fixations. *IEEE Transactions on Image Processing* 26, 9 (2017), 4446–4456.
[34] Michael F Land and Mary Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision Research* 41, 25-26 (2001), 3559–3565.
[35] Olivier Le Meur, Patrick Le Callet, and Dominique Barba. 2007. Predicting visual fixations on video based on low-level visual features. *Vision Research* 47, 19 (2007), 2483–2498.
[36] Sungkil Lee, Gerard Jounghyun Kim, and Seungmoon Choi. 2009. Real-time tracking of visually attended objects in virtual environments and its application to LOD. *IEEE Transactions on Visualization and Computer Graphics* 15, 1 (2009), 6–19.
[37] Peter McLeod, Jon Driver, and Jennie Crisp. 1988. Visual search for a conjunction of movement and form is parallel. *Nature* 332, 6160 (1988), 154.
[38] Mark Mills, Andrew Hollingworth, Stefan Van der Stigchel, Lesa Hoffman, and Michael D Dodd. 2011. Examining the influence of task set on eye movements and fixations. *Journal of Vision* 11, 8 (2011), 17–17.
[39] Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. 2011. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 103–112.
[40] Anneli Olsen. 2012. The Tobii I-VT fixation filter. *Tobii Technology* (2012).
[41] Alice J O'Toole and Candice L Walker. 1997. On the preattentive accessibility of stereoscopic disparity: Evidence from visual search. *Perception & Psychophysics* 59, 2 (1997), 202–218.
[42] Derrick Parkhurst, Klinton Law, and Ernst Niebur. 2002. Modeling the role of salience in the allocation of overt visual attention. *Vision Research* 42, 1 (2002), 107–123.
[43] Robert J Peters and Laurent Itti. 2008. Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception (TAP)* 5, 2 (2008), 9.
[44] Robert J Peters, Asha Iyer, Laurent Itti, and Christof Koch. 2005. Components of bottom-up gaze allocation in natural images. *Vision Research* 45, 18 (2005), 2397–2416.
[45] Michael J Proulx. 2007. Bottom-up guidance in visual search for conjunctions. *Journal of Experimental Psychology: Human Perception and Performance* 33, 1 (2007), 48.
[46] Michael J Proulx and Monique Green. 2011. Does apparent size capture attention in visual search? Evidence from the Müller–Lyer illusion. *Journal of Vision* 11, 13 (2011), 21–21.
[47] Constantin A Rothkopf, Dana H Ballard, and Mary M Hayhoe. 2007. Task and context determine where you look. *Journal of Vision* 7, 14 (2007), 16–16.
[48] Tim J Smith, Daniel Levin, and James E Cutting. 2012. A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science* 21, 2 (2012), 107–113.
[49] Michael J Spivey, Melinda J Tyler, Kathleen M Eberhard, and Michael K Tanenhaus. 2001. Linguistically mediated visual search. *Psychological Science* 12, 4 (2001), 282–286.
[50] Nicholas T Swafford, José A Iglesias-Guitian, Charalampos Koniaris, Bochang Moon, Darren Cosker, and Kenny Mitchell. 2016. User, metric, and computational evaluation of foveated rendering methods. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 7–14.

[51] Benjamin W Tatler, Roland J Baddeley, and Benjamin T Vincent. 2006. The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision research* 46, 12 (2006), 1857–1862.
[52] Benjamin W Tatler, Mary M Hayhoe, Michael F Land, and Dana H Ballard. 2011. Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision* 11, 5 (2011), 5–5.
[53] Anne M Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97–136.
[54] Po-He Tseng, Ran Carmi, Ian GM Cameron, Douglas P Munoz, and Laurent Itti. 2009. Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision* 9, 7 (2009), 4–4.
[55] Timothy J Vickery, Li-Wei King, and Yuhong Jiang. 2005. Setting up the target template in visual search. *Journal of Vision* 5, 1 (2005), 8–8.
[56] Martin Weier, Thorsten Roth, Ernst Kruijff, André Hinkenjann, Arsène Pérard-Gayot, Philipp Slusallek, and Yongmin Li. 2016. Foveated Real-Time Ray Tracing for Head-Mounted Displays. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 289–298.
[57] Jeremy M Wolfe, George A Alvarez, Ruth Rosenholtz, Yoana I Kuzmova, and Ashley M Sherman. 2011. Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics* 73, 6 (2011), 1650.
[58] Jeremy M Wolfe and W Gray. 2007. Guided search 4.0. *Integrated models of cognitive systems* (2007), 99–119.
[59] Jeremy M Wolfe and Todd S Horowitz. 2004. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience* 5, 6 (2004), 495.
[60] Jeremy M Wolfe and Todd S Horowitz. 2017. Five factors that guide attention in visual search. *Nature Human Behaviour* 1, 3 (2017), 0058.
[61] Jeremy M Wolfe, Aude Oliva, Todd S Horowitz, Serena J Butcher, and Aline Bompas. 2002. Segmentation of objects from backgrounds in visual search tasks. *Vision Research* 42, 28 (2002), 2985–3004.
[62] Alfred L Yarbus. 1967. Eye movements during perception of complex objects. In *Eye movements and vision*. Springer, 171–211.
[63] Qi Zhao and Christof Koch. 2011. Learning a saliency map using fixated locations in natural scenes. *Journal of Vision* 11, 3 (2011), 9–9.